Supporting Information

Lygeros *et al.* 10.1073/pnas.0805549105

SI Text

Model Description

Discrete dynamics. The discrete dynamics of the model capture instantaneous changes in the state of the system. One such change is the transition of an origin from the prereplicative to the postreplicative state. An origin is called prereplicative at the beginning of S-phase, when it is ready to fire and becomes postreplicative either when it fires or when it is passively replicated by a replication fork from an adjacent origin. Although in reality this transition takes some time, it is much faster than the rest of the dynamics of the system (e.g., the movement of the replication forks) and can therefore be accurately abstracted by an instantaneous switch that takes place when either the origin fires, or when a replication fork of an adjacent origin reaches it.

For logistical reasons, the distinction between pre- and postreplicative state needs to be refined further in our model. In particular, we need to distinguish whether the postreplicative state is reached because the origin fired or because it was passively replicated. After an origin has fired, we also need to distinguish cases where both its replication forks are active, from cases where only the right fork is active (because the left fork encountered the right replication fork of another active origin to the left of the origin in question), from cases where only the left fork is active (for the symmetric reason), from cases where neither fork is active. The discrete state of each origin, i, can be captured by a variable, S_i, that takes one of six values,

$$S_i \in \{PreR, RB, RR, LR, PostR, PassR\}.$$

Initially all origins are in the prereplicative state, therefore $S_i = PreR$ for all $i = 1, 2, \ldots, N$. The transition from *PreR* to *RB* is "spontaneous" and takes place when the origin "decides" that the time to fire has come. In our model, the timing of this transition is probabilistic; the mechanism driving it is described under the paragraph Stochastic Dynamics below. The remaining transitions are "forced," in the sense that they have to take place when certain conditions on the movement of the replication forks are met. For example, the transition of an origin i from *PreR* to *PassR* (that represents passive replication) takes place when either the left replication fork of the first active origin to the right of i, or the right replication fork of the first active origin to the left of i reaches the position, X_i, of i. States RB, LR and RR are used to discriminate origins from which active forks emanate to both directions (*RB*), only to the left (*LR*) or only to the right (*RR*) due to encountering a fork progressing from the opposite direction (fork conversion). In the conventional licensing terminology, *PassR*, *PostR*, *RB*, *RR*, and *LR* are subdivisions of the postreplicative state. Transitions between states are governed by "guards" (listed next to the arrow that represents the corresponding transition in Fig. 1 of the main text). The guards are logical statements involving the variables of the model. When the statement becomes true the corresponding transition is taken.

The guards make use of the right and left neighbors of a given origin, i.e., origins to the left and to the right that are actively replicating at this instant

$$LN(i) = \max\{j < i | S_j \notin \{PreR, PostR, PassR\}\}$$
$$RN(i) = \min\{j > i | S_j \notin \{PreR, PostR, PassR\}\}$$

The guard for the transition from *PreR* to *PassR* can then be succinctly written as

$$G_{PreR \to PassR} = [X_{LN(i)} + R_{LN(i)} \geq X_i] \vee [X_{RN(i)} - L_{RN(i)} \leq X_i],$$

where $R_{LN(i)}$ and $L_{RN(i)}$ denote the progress of the right and left replication forks (respectively) of these neighboring origins, and $\vee$ denotes "or." The remaining transition guards are

$$G_{RB \to RR} = [X_{LN(i)} + R_{LN(i)} \geq X_i - L_i]$$
$$G_{PreR \to RB} = [t \geq T_i]$$
$$G_{RB \to LR} = [X_{NN(i)} - L_{RN(i)} \leq X_i + R_i]$$
$$G_{RR \to PostR} = G_{RB \to LR}$$
$$G_{LR \to PostR} = G_{RB \to RR}$$

The guard $G_{PreR \to RB}$ reflects the probabilistic transition that takes place when origin i fires; t denotes the current time and T_i the random origin firing time, whose probability distribution is discussed below.

Continuous dynamics. The continuous dynamics of the model capture evolutions that are slow compared to the discrete transitions discussed above and can therefore be considered as taking place continuously over a time interval. In our model the only such evolution is the movement of the replication forks. The progress of the replication forks is measured in terms of the number of bases replicated. Even though this number is finite by nature, it is very large and it is therefore reasonably accurate to capture the progress of the replication fork by a continuous quantity that changes according to a differential equation. The rationale behind this approximation becomes clearer if we consider normalizing the number of replicated bases by the total number of bases, L, (which is more than 12 million for *S. pombe*) and looking at the resulting number, which will be between *0* and *1* (reflecting the fraction of the genome that has been replicated).

When origin i fires it gives rise to two replication forks moving away from the origin to the left and to the right. We denote by L_i and R_i the number of bases that these forks have replicated respectively. The forks move (i.e., L_i, and R_i increase) at a velocity $v(x)$ which depends on the position, x, of the genome currently being replicated by the fork. The progress of the forks stops when they encounter replication forks moving in the opposite direction (reflected in the discrete state of the model). The progress of the replication forks can therefore be captured by two differential equations

$$\frac{dR_i(t)}{dt} = \begin{cases} v(X_i + R_i(t)) & \text{if } S_i(t) \in \{RB, RR\} \\ 0 & \text{otherwise} \end{cases}$$

$$\frac{dL_i(t)}{dt} = \begin{cases} v(X_i - L_i(t)) & \text{if } S_i(t) \in \{RB, LR\} \\ 0 & \text{otherwise} \end{cases}$$

Clearly, the moment origin i fires the number of bases its forks have replicated is zero, therefore we initialize the differential equations with $L_i(0) = R_i(0) = 0$. Notice that there is a tight coupling between the discrete and continuous states of the process: The transitions of the discrete state depend on the value of the continuous state and the evolution of the continuous state depends on the value of the discrete state. This tight coupling of

the continuous and the discrete is the defining feature of hybrid systems.

Stochastic dynamics. In our model we assume that, in the absence of passive replication, the firing time, T_i (in minutes), of origin i is governed by an exponential distribution. More specifically, we assume that in the absence of passive replication the probability that origin i has not fired by time t (in minutes) decreases exponentially in t:

$$\text{Probability}[T_i \geq t] = e^{-\lambda_i t}.$$

Recall that the positive number λ_i reflects the firing propensity of origin i in a unit of time in the absence of passive replication. The exponential distribution offers theoretical and practical advantages and is widely used in many applications, among them telecommunication networks, manufacturing, insurance, and finance. For example, the exponential distribution is memoryless. Moreover, the parameter λ_i may in general depend on the discrete state of the model (e.g., the number of origins that are prereplicative) and/or the continuous state (e.g., the fraction of unreplicated DNA). This observation is exploited below for the development and simulation of the firing propensity redistribution model.

The parameter λ_i for each origin needs to be selected based on experimental data. To do this we use the fraction of cells, FP_i, in which origin i was observed to fire in hydroxyurea experiments, where replication forks are halted after moving only for a few thousand bases, thus preventing passive replication. It is easy to see that under the exponential distribution and in the absence of passive replication the probability that origin i fires by a given time T_f (in minutes) is given by

$$\int_0^{T_f} \lambda_i e^{-\lambda_i t}\, dt = 1 - e^{-\lambda_i T_f}.$$

If we equate this number to FP_i we can obtain an estimate of the intrinsic firing propensity, λ_i, of origin i as a function of the experimentally observed fraction FP_i

$$\lambda_i = -\frac{1}{T_f}\ln(1 - FP_i).$$

Clearly there is some arbitrariness involved in the choice of T_f. Initially we set $T_f = 20$ minutes to reflect the generally accepted view that the S-phase for *S pombe* lasts about 20 min. The sensitivity of the results with respect to this choice is discussed in the main body of the paper.

We note two subtle assumptions of our model. The first is that the firing processes of different origins are assumed to be statistically independent. The second is that the value of λ_i is assumed to be constant during S-phase. This implicitly means that origin efficiency (weak *vs.* strong) and origin firing time (early *vs.* late) both boil down to the same mechanism; strong origins will also tend to fire early and weak origins will tend to fire late. The reason is that the mean of the exponential distribution is $1/\lambda_i$. Therefore, a high value of λ_i implies that, in the absence of passive replication, origin i will tend to fire both more often and earlier. In the presence of passive replication the correlation will not be perfect, of course. An alternative firing propensity re-distribution model that relaxes both assumptions is discussed below.

Firing propensity redistribution model. The structure of the firing propensity redistribution model is the same as that of the basic model outlined above. The only difference lies in the calculation of the intrinsic firing propensities which change every time an origin fires or gets passively replicated. As before, let λ_i denote the initial firing propensity of origin i, computed by the method discussed in the previous section and let $\bar{\lambda}_i(t)$ denote the firing propensity of origin i at time t; initially we set $\bar{\lambda}_i(0) = \lambda_i$. Let also

$$Q(t) = \{i | S_i = PreR\} \subseteq \{1,2,\ldots,N\}$$

denote the set of indices of the origins in the prereplicative state at time t; initially we set $Q(0) = \{1, 2, \ldots, N\}$. Assume that at some time, t, origin i either fires, or becomes passively replicated. At this time, the index i is dropped from the set $Q(t)$, and the firing propensity of origin i is redistributed among the origins remaining in $Q(t)$. We assume that the redistribution is proportional to the initial firing propensity of the remaining prereplicative origins, in other words

$Q(t) = Q(t^-) - \{i\}$ and

$$\bar{\lambda}_j(t) = \bar{\lambda}_j(t^-) + \bar{\lambda}_i(t^-)\frac{\lambda_j}{\sum_{k\in Q(t)}\lambda_k} \text{ for all } j \in Q(t)$$

where t^- denotes the time just before the discrete transition (firing or passive replication) takes place; more formally, $Q(t^-) = \lim_{\tau \nearrow t} Q(\tau)$. It is easy to see that in this case the firing propensity of individual origins increases monotonically along the S-phase, while the total firing propensity remains constant

$$\bar{\lambda}(t) = \sum_{i\in Q(t)} \bar{\lambda}_i(t) = \sum_{i=1}^{N} \lambda_i$$

Two alternative redistribution models were also implemented. The first alternative is the same as above, with the exception that redistribution takes place according to the current firing propensity of the *PreR* origins and not their initial firing propensity. In other words

$$\bar{\lambda}_j(t) = \bar{\lambda}_j(t^-) + \bar{\lambda}_i(t^-)\frac{\bar{\lambda}_j(t^-)}{\sum_{k\in Q(t)}\bar{\lambda}_k(t^-)} \text{ for all } j \in Q(t)$$

For this first alternative redistribution model the total firing propensity also remains constant.

The second alternative redistribution model captures the situation where redistribution takes place upon fork conversion and not upon firing. In this case, redistribution takes place either when origin i is passively replicated (transition *PreR* to *PassR*) or when forks meet (transitions *RB* to *RR*, *RB* to *LR*, *RR* to *PostR*, or *LR* to *PostR*). In both cases, the computation of $Q(t)$ proceeds as before. In the former case all of the firing propensity $\bar{\lambda}_i(t^-)$ gets redistributed among origins in $Q(t)$ proportionately to their initial firing propensities. In the latter case, one half of the firing propensity $\bar{\lambda}_i(t^-)/2$ gets redistributed. One can see that in this case the total firing propensity will not be exactly constant, since origins in the *RB*, *RR*, and *LR* states are not in $Q(t)$ but still hold on to (part of) their firing propensity.

In all cases, the parameters used in the results [L, N, T_f, $v(x)$] were the same as for the base case of the original model.

Fission yeast instantiation: Modeled genome areas and input parameters. The locations and intrinsic firing propensities of origins along the fission yeast genome were taken from (Heichinger *et al.*, 2006). The genome sequence release of the fission yeast genome of April 2004 from the Sanger Center was used. Highly repetitive regions of the fission yeast genome were excluded from the origin mapping analysis and were therefore also excluded from simulations. These are telomeric and subtelomeric repeats, centromeric and subcentromeric repeats, and ribosomal RNA repeats. The left and right arms of chromosomes 2 and 3 were modeled as separate pieces to avoid artifacts from artificially

joining them following extraction of centromeric regions. Chromosome 1 was modeled as one piece, since the length of the unmapped centromeric region was ≈15 kb, which is no different from the mean inter-origin distance. DNA replication was modeled for a total area of $L = 12{,}039{,}987$ bases containing $N = 893$ potential origins. Specifically the modeled regions were:

- Chromosome 1: base 90,532 to 5,494,088. 5.4 Mb containing 408 origins
- Chromosome 2-Left arm: base 95,383 to 1,592,054. 1.5 Mb containing 105 origins.
- Chromosome 2-Right arm: base 1,649,764 to 4,426,827. 2.8 Mb containing 207 origins.
- Chromosome 3-Left arm: base 33,072 to 1,065,271. 1 Mb containing 83 origins.
- Chromosome 3-Right arm: base 1,144,065 to 2,431,896. 1.4 Mb containing 90 origins.

Location and intrinsic firing propensities of origins used as model input are shown in supporting information (SI) Table S1.

To specify the locations of additional putative origins, a bioinformatics analysis of the properties of all known fission yeast origins was carried out. Consistent with previous analyses, mapped fission yeast origins were shown to localize to intergenic regions which were over 500 bp in length. The maximum AT content of 500 bp windows within each intergenic region was shown to be a good predictor of origin activity. A moving AT content threshold was therefore used to progressively increase the number of intergenic regions that were included as weak origins.

Hardware and software. The model was implemented in Matlab 2006b on a dual core (2 × 1,83GHz) computer with 2048MB of RAM. The results are generally based on 2,000 Monte-Carlo runs. To keep the memory requirements manageable, continuous states were added as they became necessary and dropped after they ceased being active. Forced transitions (all except the transition from *PreR* to *RB*) were simulated by the event detection facilities of the ode45 Matlab routine used to integrate the continuous dynamics. The spontaneous, probabilistic transition from *PreR* to *RB* was simulated by extracting a random variable uniformly in the unit interval, taking its logarithm, dividing by $-\lambda_i$, and waiting for the simulation time to reach the resulting value. The probability distribution for the time at which this happens obeys the exponential distribution with rate λ_i.

Simulation and diagnostics. A simple calculation suggests that, even though the model used to capture the behavior of individual origins is relatively simple, the resulting genome-wide model can be very complex. The number of discrete states and the dimension of the continuous state space can be very large, with up to $2N$ continuous states and 6^N discrete states being activated, with $N = 893$ for the fission yeast genome. While only a fraction of these states will be visited in any one execution of the model, the analysis and even the simulation of such a model can be a formidable task. In the implementation of our model for computer simulation special care had to be taken to ensure that the simulations were faithful to the model and simulation time and memory requirements were kept reasonable.

Fig. S1 shows examples of simulations that represent full genome replication in individual cells for the basic model. In Fig. S1*A*, the replication process in a specific region of the genome, containing 11 origins, is depicted for two simulations (C_1 and C_2). The stochastic nature of the process is evident in the different location of active origins in each simulation, the different timing of firing of each origin and the different total time required to complete replication of this genomic region. In Fig. S1*B*, the replication time of each position x along a second genomic region is shown for four simulations. The stochasticity of the replication process is evident.

To ensure that the programming code accurately captures the model dynamics, two separate implementations were generated. The first was based on a continuous time simulation of the differential equations governing the evolution of the continuous state, coupled with event detection to simulate the evolution of the discrete state. The second implementation is based on an explicit algebraic solution of the differential equations, followed by a discrete time simulation from one discrete event to the next. The former implementation is more general but also more computationally demanding, while the latter only applies to cases where the fork speed is constant, but can be considerably faster. Both implementations were tested on the fission yeast instantiation discussed here and produced identical results (albeit at a different computational cost). This is a clear indication that both computational implementations indeed capture the dynamics of the mathematical model correctly.

Next, to ensure that the model dynamics correctly capture our current understanding of the DNA replication process, several tests were carried out using the two computational implementations. The simulation tests showed the following properties:

1. The fraction of unreplicated DNA always starts at 1 and decays monotonically to 0.
2. The total amount of DNA in the nucleus starts with 1 genome length and monotonically increases to 2 genome lengths.
3. All parts of the genome (at inter-putative origin granularity) started with one copy and ended with two.
4. All origins started in the *PreR* state and ended up either in the *PostR* state or in the *PassR* state. Further, the number of *PreR* origins was monotone decreasing while the number of *PostR* and *PassR* origins was monotone increasing.
5. The number of *RR* origins was always (roughly) equal to the number of *LR* origins. Small differences are due to origins near the edges of the chromosome pieces completing the replication to the end of the piece.

Fig. S2 demonstrates these properties for two runs of the model, one for the base-case model (without redistribution, left column) and one for the firing propensity redistribution model (right column).

1. Heichinger C, Penkett CJ, Bahler J, Nurse P (2006) *EMBO J* 25:5171–9.

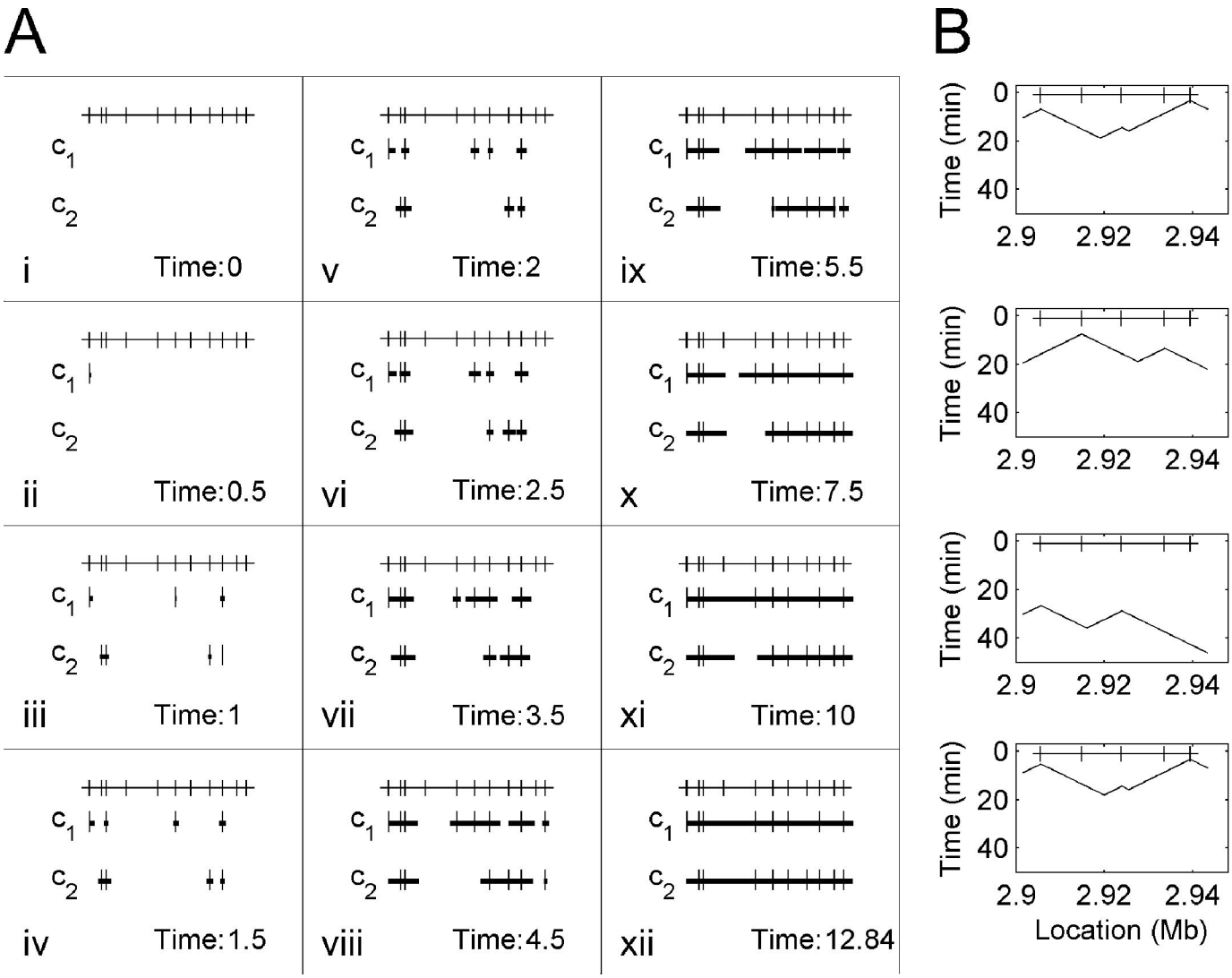

Fig. S1. Example simulation runs of DNA replication across the fission yeast genome. (*A*) Snapshots of a ''movie'' of two simulations of the replication process in a given piece of the genome containing 11 putative origins. For each snapshot the bar at the top shows the locations of potential origins, the two bars labeled C_1 and C_2 show the two simulations, and the time corresponding to the snapshot is listed in the bottom; origins that have fired in the simulations are marked by a vertical line and the replicated pieces of the DNA by a thick horizontal line. Stochastic phenomena in the location of active origins and timing of firing are evident. As a consequence, the time required to complete replication of this region differs between simulations (10 min for C_1 verus 13 min for C_2). (*B*) Results of 4 different simulations a 40-kb piece of the genome containing 5 putative origins, showing the time of replication of each genomic region. Locations of potential origins are marked at the top, the horizontal axis represents location in the genome and the vertical axis the time at which each location was replicated in the particular simulation. Differences in the timing profile in different simulations are evident.

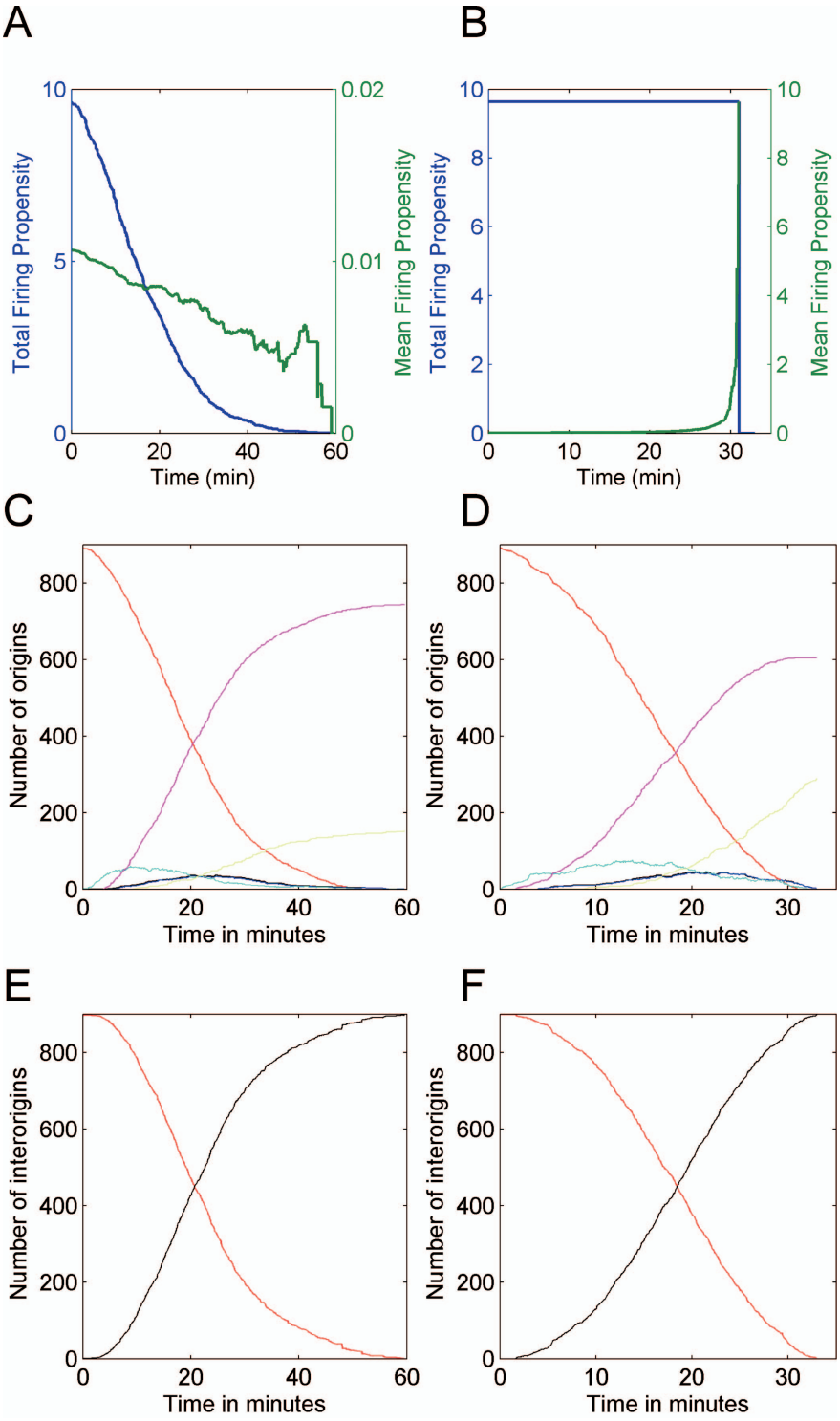

Fig. S2. Properties of base-case and firing propensity redistribution model. (*A*, *C*, and *E*) Base case model, (*B*, *D*, and *F*) Firing propensity redistribution model. (*A*) and (*B*) show the total system firing propensity (the sum of the firing propensities of origins available to fire, in blue) and the mean firing propensity of origins available to fire (in green) during the course of S-phase. Note that the total system firing propensity decreases during the course of S-phase for the base-case model and remains constant for the firing propensity redistribution model. In contrast, the firing propensity of a given origin remains constant in the base case model and increases in the redistribution model. The mean firing propensity tends to decrease somewhat in the base model, as more efficient origins tend to fire earlier. (*C*) and (*D*) show the number of origins that are found in each one of the six discrete states during the course of S-phase. *PreR* red, *PassR* purple, *RB* light blue, *RL* dark blue *RR* black, and *PassR* yellow. Note that the RL and RR curves effectively overlap and that for both models, the majority of origins are passively replicated. (*E*) and (*F*) show the number of inter-origin locations which have one copy (black line) or two copies (red line) during the course of S-phase.

Other Supporting Information Files

Table S1.xls